Predicting Flight Delays Using Weather

December 12, 2024
Team Data Hawks

Ella Bronaugh, Mayur Dalvi, Reza Naiman, Tristan Levy-Park

Research Question:

Are weather conditions an effective predictor in classifying flight delays?

Data Analysis Timeline

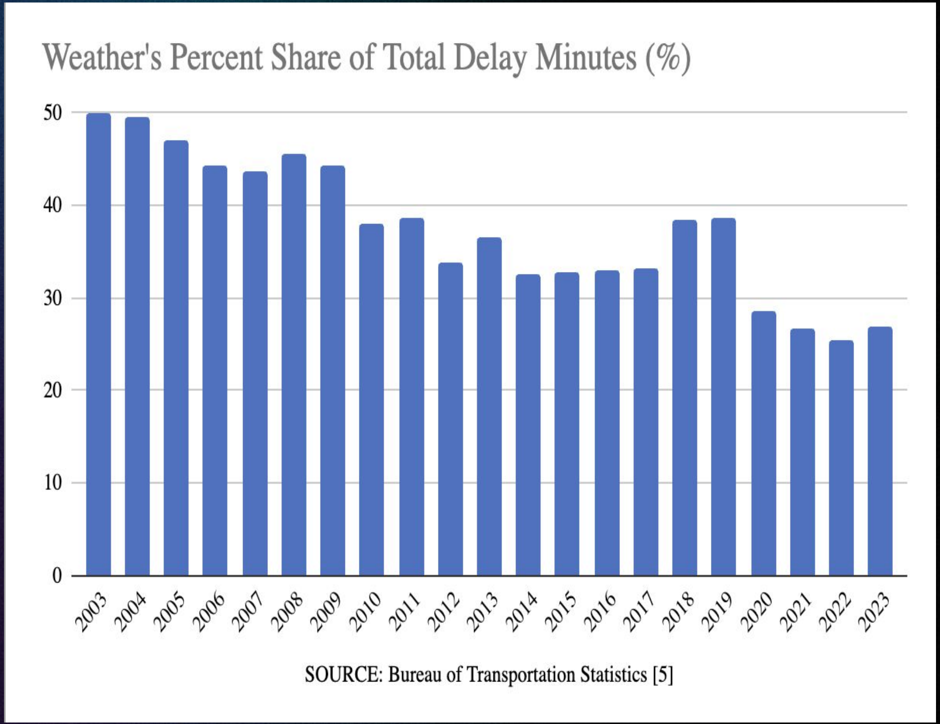
- Research and Lit Review
- Data Collection and Cleaning
- Modeling
- Results and Discussion

Impact of Weather-Related Flight Delays

1. **Delays:** Extreme weather caused 32.6% of delay minutes (2003-2015), with up to 82% in severe cases. [2]

2. **Economic Costs:** Delays cost up to $40.2B annually, including $31.2B in 2010. [3]

3. **Environmental Impact:** Idling/rerouting increase emissions and air pollution. [4]



Data Collection

Source:

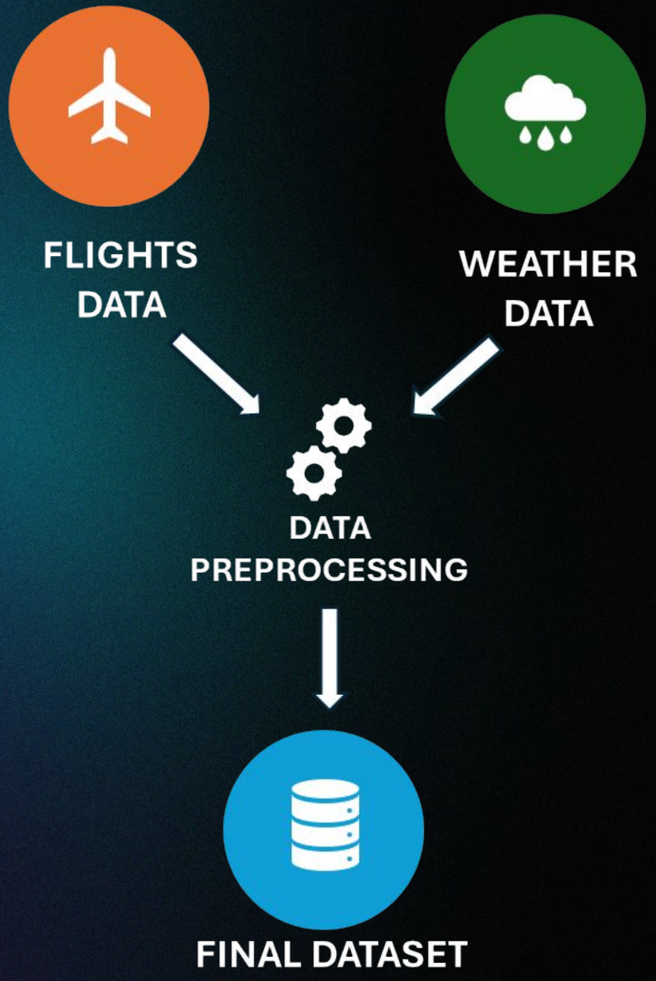
- Flight Data : Bureau of Transportation Statistics
- Weather Data : Iowa Environmental Mesonet

Merging Conditions:

- Columns: DateTime & Location

DataSet information:

- Flight Features: 18
- Weather Features: 12
- States: Illinois & Georgia (10 years)
- Size: 1 GB



Data Cleaning & Preprocessing

Flights Data:

- Dropped missing data row
- Feature-engineered column to create datetime feature for merging condition
- Most of the columns were strings that had to be converted into categories.
- Flight delayed more than 15 min it is classified as delayed flights.

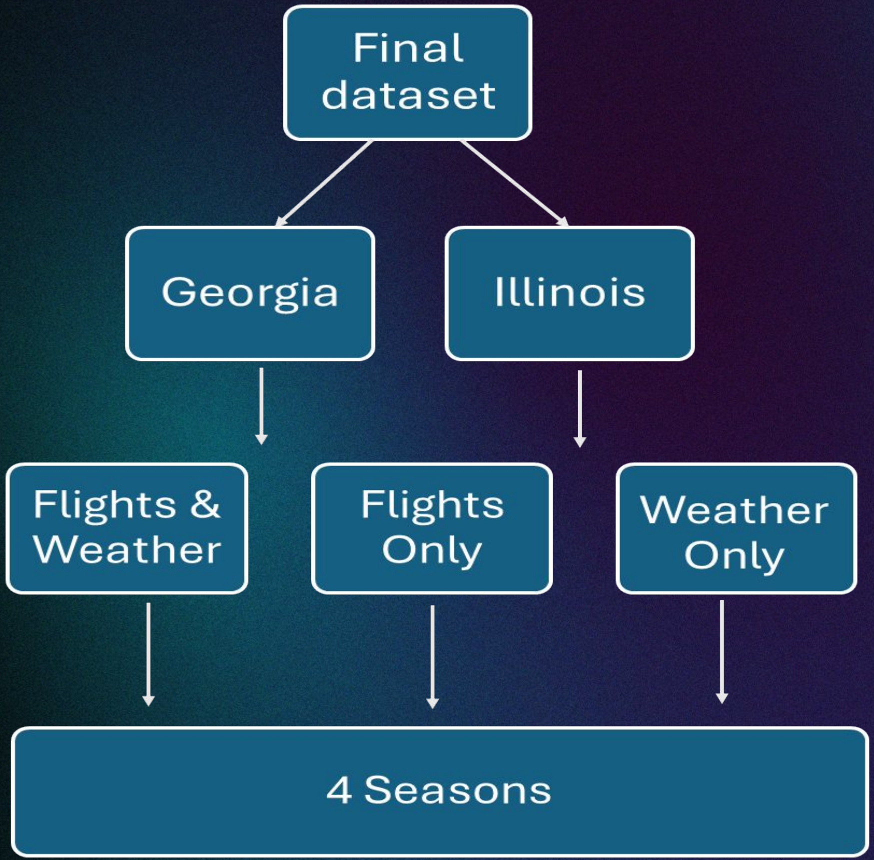
Weather Data:

- Where there was ‘M‘ replaced with NaN
- Applied forward fill to missing data
- Generated a 15-min interval and filled it with linear interpolation
- For precipitation where there was ‘T’ it was replaced with 0.005.

Methodology and Modeling

- SMOTE for Balancing: Applied SMOTE to address class imbalance in the dataset.
- Data Split (70-20-10): 10-year data split: 7 years for training, 2 for validation, 1 for testing.
- Model Exploration: Evaluated 5 models: Logistic Regression, Linear SVC, Random Forest, Decision Tree, CatBoost.

Three Fold Method, Months, and Seasons



Results & Discussion

Testing Set (IL Dataset):

- Using **Weather Information**: 70.6% accuracy
- Using **Flight Information**: 76.8% accuracy
- Using **Combination of Both**: 73.6% accuracy

Testing Set (GA Dataset):

- Using **Weather Information**: 48.75% accuracy
- Using **Flight Information**: 64.5% accuracy
- Using **Combination of Both**: 71.3% accuracy

Feature Importance

Weather data is an effective predictor for flight delays sometimes.

Illinois, January:

Weather Predictors Only:

Model	Accuracy	Precision	Recall	F1-Score
Full LR	0.64	0.41	0.65	0.51
Linear SVC	0.65	0.42	0.64	0.51
DT	0.76	0.77	0.19	0.31
RF	0.57	0.34	0.60	0.44
CatBoost	0.41	0.30	0.87	0.45

Flight Predictors Only:

Model	Accuracy	Precision	Recall	F1-Score
Full LR	0.56	0.31	0.46	0.37
Linear SVC	0.71	0.15	0.00	0.01
DT	0.77	0.7	0.31	0.43
RF	0.64	0.34	0.30	0.32
CatBoost	0.46	0.32	0.81	0.46

Flight + Weather Predictors:

Model	Accuracy	Precision	Recall	F1-Score
Full LR	0.64	0.41	0.65	0.50
Linear SVC	0.75	0.71	0.16	0.26
DT	0.77	0.70	0.31	0.43
RF	0.56	0.36	0.73	0.48
CatBoost	0.66	0.42	0.53	0.47

Illinois, April:

Weather Predictors Only:

Model	Accuracy	Precision	Recall	F1-Score
Full LR	0.35	0.25	0.85	0.38
Linear SVC	0.28	0.23	0.90	0.37
DT				
RF	0.33	0.24	0.86	0.38
CatBoost	0.23	0.23	0.99	0.38

Flight Predictors Only:

Flight + Weather Predictors:

Illinois, July:

Weather Predictors Only

Flight Predictors Only

Flight + Weather Predictors

Illinois, October:

Weather Predictors Only

Flight Predictors Only

Flight + Weather Predictors

Georgia, January:

Weather Predictors Only

Flight Predictors Only

Flight + Weather Predictors

Georgia, April:

Weather Predictors Only

Flight Predictors Only

Flight + Weather Predictors

Georgia, July:

Weather Predictors Only

Flight Predictors Only

Flight + Weather Predictors

Georgia, October:

Weather Predictors Only

Flight Predictors Only

Flight + Weather Predictors

Paper Formatting and Influence

Research done by Kim & Park heavily influenced the format of our paper [1]

Layout and Description of Features:

Attribute mame	Description	Mean (Std)	Min	Max
Time (year)	2010–2021 (e.g. 2020)	–	–	–
Airline	Unique carrier [e.g. AA (American Airlines)]	–	–	–
Flight number	Flight number (e.g. AA2000)	–	–	–
Destination	Destination (e.g. JFK)	–	–	–
Planned departure time	Planned departure time (e.g. 1622)	–	–	–
Actual departure time	Actual departure time (e.g. 1634)	–	–	–
Result status	Takeoff intime or delay status (e.g. 1)	–	–	–
Delay type	Delay type (e.g. WeatherDelay)	–	–	–
Wind direction	Wind direction (e.g. NW, WNW)	–	–	–
Wind speed	Wind speed (e.g. 3)	10.5 (5.3)	0	51
Wind gust	Wind gust (e.g. 24)	5.3 (10.9)	0	75
Temperature (celcius)	Temperature (celcius) (e.g. 34)	51.5 (20.5)	−21	103
Dew point temperature (celcius)	Dew point temperature (celcius) (e.g. 31)	39.9 (19.5)	−32	79
Humidity	Humidity (e.g. 92)	67.7 (17.2)	0	100
Pressure (hPa)	Pressure (hPa) (e.g. 29.96)	29.3 (0.3)	0	30.2
Precipitation (mm)	Precipitation (mm) (e.g. 0.1)	0.006 (0.046)	0	2
Condition	Condition (e.g. Cloudy, Windy)	–	–	–

Layout of Important Metrics:

Algorithm		Time difference: 2 h					
		Accuracy	Precision	Recall	F1-score	Train (s)	Test (us)
DT	Normal	0.688	0.704	0.676	0.690	0.112	0.318
	Delayed		0.671	**0.700**	0.685		
RF	Normal	**0.749**	**0.729**	**0.814**	**0.769**	2.254	16.242
	Delayed		**0.776**	0.680	**0.725**		
SVM	Normal	0.651	0.631	0.774	0.695	3.625	458.280
	Delayed		0.686	0.522	0.593		
KNN	Normal	0.641	0.655	0.637	0.646	0.003	60.510
	Delayed		0.628	0.646	0.637		
LR	Normal	0.595	0.600	0.635	0.617	0.085	0.318
	Delayed		0.589	0.552	0.570		
XGB	Normal	0.721	0.715	0.759	0.736	0.150	1.274
	Delayed		0.728	0.680	0.703		
LSTM	Normal	0.644	0.620	0.776	0.689	490.4	3.503
	Delayed		0.687	0.509	0.584		

References

[1] Kim, S., Park, E. Prediction of flight departure delays caused by weather conditions adopting data-driven approaches. J Big Data 11, 11 (2024).

[2] Goodman, C. J., and J. D. Small Griswold, 2019: Meteorological Impacts on Commercial Aviation Delays and Cancellations in the Continental United States. J. Appl. Meteor. Climatol., 58, 479–494

[3] Yhdego et al. "Analyzing the Impacts of Inbound Flight Delay Trends on Departure Delays Due to Connection Passengers Using a Hybrid RNN Model." Mathematics 2023, 11, 2427.

[4] Gratton, G. B. et al. "Reviewing the Impacts of Climate Change on Air Transport Operations." The Aeronautical Journal 126.1295 (2022): 209–221. Web.

[5] https://www.bts.gov/content/weathers-share-delay-percent-total-delay-minutes-year